SPECIFICATION

SOUND ACQUISITION METHOD AND SOUND ACQUISITION APPARATUS

TECHNICAL FIELD

The present invention relates to a sound acquisition method and a sound acquisition apparatus and, more particularly, to a sound acquisition method and a sound acquisition apparatus that acquire speech sounds from a plurality of speech sound sources and adjust their volume before outputting.

PRIOR ART

For example, in a teleconference in which persons at different remote locations participate, if a single microphone is used at each location to acquire speech sounds of plural participants sitting at different positions at each remote location, received signal levels greatly differ because of different distances from the participants to the microphone and different volumes of their speech sounds. At the remote receiving side the reproduced speech sounds greatly differ in volume with the participants at the transmitting side and, in some cases, they are hard to distinguish one participant from another.

Fig. 17 illustrates in block form the basic configuration of a conventional sound acquisition apparatus disclosed, for example, in Japanese Patent Application Kokai Publication 8-250944. The conventional sound acquisition apparatus is made up of a microphone 41, a power calculating part 42, an amplification factor setting part 43, and an amplifier 44. The power calculating part 42 calculates a long-time mean power P_{ave} of the signal received by the microphone 41. The long-time mean power can be obtained by squaring the signal and time-integrating the squared output. Next, the

amplification factor setting part 43 sets an amplification factor G based on the long-time mean power P_{ave} of the received signal calculated by the power calculating part 42 and a preset desired sending level P_{opt}. The amplification factor G can be calculated, for example, by the following equation (1).

$$G = (P_{opt}/P_{ave})^{1/2} \quad (1)$$

The amplifier 44 amplifies the microphone received signal by the set amplification factor G and outputs the amplified signal.

By the processing described above, the output signal power is put to the desired sending level P_{opt}, by which the volume is automatically adjusted. With the conventional sound acquisition method, however, since the amplification factor is determined based on the long-time mean power, a delay of several to tens of seconds develops in the setting of an appropriate amplification factor. Accordingly, in the case where plural speakers are present and their speech sounds are acquired by the microphone at different levels, there arises a problem that whenever the speakers changes from one to another, setting of an appropriate amplification factor delays, resulting in an the speech sound being reproduced at an inappropriate volume.

An object of the present invention is to provide a sound acquisition apparatus and a sound acquisition method that, even where plural speakers are present and their speech sounds are picked up by a microphone at different levels, automatically adjust the volume of each speech sound to an appropriate value, and a program for implementing the method.

DISCLOSURE OF THE INVENTION

A sound acquisition method for acquiring sound from each sound source by microphones of plural channels according to the present invention, comprises:

(a) a state deciding step including an utterance deciding step of deciding an utterance period from signals received by said plural-channel microphones;

(b) a sound source position detecting step of detecting the position of said each sound source from said received signals when the utterance period is decided in said utterance deciding step;

(c) a frequency domain converting step of converting said received signals to frequency domain signals;

(d) a covariance matrix calculating step of calculating a covariance matrix of said frequency domain received signals;

(e) a covariance matrix storage step of storing said covariance matrix for each sound source based on the result of detection in said sound position detecting step;

(f) a filter coefficient calculating step of calculating filter coefficients of said plural channels based on said stored covariance matrix and a predetermined output level;

(g) a filtering step of filtering the received signals of said plural channels by filter coefficients of said plural channels, respectively; and

(h) an adding step of adding together the results of filtering in said plural channels, and providing the added output as a send signal.

According to the present invention, a sound acquisition apparatus which acquires sound from each sound source by microphones of plural channels placed in an acoustic space, comprises:

a state decision part including an utterance deciding part for deciding an utterance period from signals received by said plural-channel microphones;

a sound source position detecting part for detecting the position of said

each sound source from said received signals when the utterance period is decided by said utterance deciding part;

a frequency domain converting part for converting said received signals to frequency domain signals;

a covariance matrix calculating part for calculating a covariance matrix of said frequency domain received signals of said plural channels;

a covariance matrix storage part for storing said covariance matrix for said each sound source based on the result of detection by said sound position detecting part;

a filter coefficient calculating part for calculating filter coefficients of said plural channels by use of said stored covariance matrix so that the send signal level for said each sound source becomes a desired level;

filters of said plural channels for filtering the received signals from said microphones by use of the filter coefficients of said plural channels, respectively; and

an adder for adding together the outputs from said filters of said plural channels and for providing the added output as a send signal.

According to a second aspect of the present invention, a sound acquisition method for acquiring speech sound from at least one sound source by a microphone of at least one channel in an acoustic space in which a received signal is reproduced by a loudspeaker, comprises:

(a) a state deciding step of deciding an utterance period and a receiving period from the sound acquired by said microphone of said at lest one channel and said received signal;

(b) a frequency domain converting step of converting said acquired signal and said received signal to frequency domain signals;

(c) a covariance matrix calculating step of calculating a covariance

matrix in said utterance period and a covariance in said receiving period from said frequency domain acquired signal and received signal;

(d) a covariance matrix storage step of storing said covariance matrices for said utterance period and for said receiving period, respectively;

(e) a filter coefficient calculating step of calculating filter coefficients for said acquired signal of said at least one channel and filter coefficients for said received signal based on said stored covariance matrices in said utterance period and said receiving period so that an acoustic echo, which is a received signal component contained in said received signal is cancelled;

(f) a filtering step of filtering said received signal and said acquired signal by use of said filter coefficients for said received signal and filter coefficients for said acquired signal of said at least one channel; and

(g) an adding step of adding together said filtered signals and providing the added output as a send signal.

A sound acquisition apparatus according to the second aspect of the present invention comprises:

a microphone of at least one channel for acquiring speech sound from a sound source and for outputting an acquired signal;

a loudspeaker for reproducing a received signal;

a state decision part for deciding an utterance period and a receiving period from said acquired signal and said received signal;

a frequency domain converting part for converting said acquired signal and said received signal to frequency domain signals;

a covariance matrix calculating part for calculating covariance matrices of said acquired and received signals of said frequency domain for said utterance period and for said receiving period, respectively;

a covariance matrix storage part for storing said covariance matrices

for said utterance period and for said receiving period, respectively;

a filter coefficient calculating part for calculating filter coefficients for said acquired signal of said at least one channel and filter coefficients for said received signal based on said stored covariance matrices so that an acoustic echo of said received signal is cancelled;

an acquired signal filter and a received signal filter having set therein said filter coefficients for said acquired signal and said filter coefficients for said received signal, for filtering said acquired signal and for filtering said received signal; and

an adder for adding together the outputs from said acquired signal filter and said received signal filter, and for providing the added output as a send signal.

According to the present invention, even when plural speakers are present ad their speech sounds are acquired by a plurality of microphones at different levels, the directivity of the microphones is appropriately controlled to automatically adjust the volume of the speech sound to an appropriate value for each speaker.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram illustrating a sound acquisition apparatus according to a first embodiment of the present invention.

Fig. 2 is a block diagram showing an example of the configuration of a state decision part 14 in Fig. 1.

Fig. 3 is a block diagram showing an example of the configuration of a sound source position detecting part 15 in Fig. 1.

Fig. 4 is a block diagram showing an example of the configuration of a filter coefficient calculating part 21 in Fig. 1.

Fig. 5 is a flowchart showing a first example of a sound acquisition method using the sound acquisition apparatus of Fig. 1.

Fig. 6 is a flowchart showing a second example of a sound acquisition method using the sound acquisition apparatus of Fig. 1.

Fig. 7 is a flowchart showing a third example of a sound acquisition method using the sound acquisition apparatus of Fig. 1.

Fig. 8 is a block diagram illustrating a sound acquisition apparatus according to a second embodiment of the present invention.

Fig. 9 is a block diagram showing an example of the configuration of the state decision part 14 in Fig. 8.

Fig. 10 is a block diagram illustrating a sound acquisition apparatus according to a third embodiment of the present invention.

Fig. 11 is a block diagram showing an example of the configuration of the state decision part 14 in Fig. 7.

Fig. 12 is a block diagram illustrating a sound acquisition apparatus according to a fourth embodiment of the present invention.

Fig. 13 is a block diagram illustrating a sound acquisition apparatus according to a fifth embodiment of the present invention.

Fig. 14 is a block diagram showing an example of the configuration of a weighting factor setting part 21H in Fig. 4.

Fig. 15 is a block diagram showing another example of the configuration of a weighting factor setting part 21H in Fig. 4.

Fig. 16 is a block diagram showing an example of the configuration of a whitening part 21J in Fig. 4.

Fig. 17 is a block diagram an example of a covariance matrix storage part 18 that is used when each embodiment is equipped with a covariance matrix averaging function.

Fig. 18A is a diagram showing simulated speech waveforms of speakers A and B before processing by the first embodiment.

Fig. 18B is a diagram showing simulated speech waveforms of speakers A and B after processing by the first embodiment.

Fig. 19 is a diagram showing received and send speech waveform by simulation, which show acoustic echo and noise cancellation according to a third embodiment.

Fig. 20 is a block diagram illustrating a conventional sound acquisition apparatus.

BEST MODE FOR CARRYING OUT THE INVENTION

FIRST EMBODIMENT

Fig. 1 is a block diagram of a sound acquisition apparatus according to a first embodiment of the present invention.

The sound acquisition apparatus of this embodiment comprises microphones 11_1 to 11_M of M channels disposed in an acoustic space, filters 12_1 to 12_M, an adder 13, a state decision part 14, a sound source position detecting part 15, a frequency domain converting part 16, a covariance matrix calculating part 17, a covariance matrix storage part 18, an acquired sound level estimating part 19, and a filter coefficient calculating part 21.

In this embodiment, the positions of speech sound sources 9_1 to 9_K in an acoustic space are detected, then covariance matrices of acquired signals in the frequency domain for the respective speech sound sources are calculated and stored, and these covariance matrices are used to calculate filter coefficients. These filter coefficients are used to filter the signals acquired by the microphones, thereby controlling the signals from the respective speech sound sources to have a constant volume. In this embodiment, let it

be assumed that the output signals from the microphones 11_1 to 11_M are digital signals into which the signals acquired by the microphones are converted by a digital-to-analog converter at a predetermined sampling frequency, though not shown in particular. This applies to other embodiments of the present invention.

In the first place, the state decision part 14 detects an utterance period from each of the received signals by the microphones 11_1 to 11_M. For example, as shown in Fig. 2, in the state decision part 14 all the received signals from the microphones 11_1 to 11_M are added together by an adding part 14A, then the added output is applied to a short-time mean power calculating part 14B and a long-time mean power calculating part 14C to obtain a shirt-time mean power (approximately in the range of 0. 1 to 1 s, for instance) P_{avS} and a long-time mean power (approximately in the range of 1 to 100 s, for instance) P_{avL}, respectively, then the ratio between the short-time mean power and the long-time mean power, $R_p=P_{avS}/P_{avL}$, is calculated in a division part 14D, and in an utterance decision part 14E the power ratio R_p is compared with a predetermined utterance threshold value R_{thU}; if the power ratio exceeds the threshold value, then the former is decided as indicating the utterance period.

When the decision result by the state decision part 14 is the utterance period, the sound source position detecting part 15 estimates the position of the sound source. A method for estimating the sound source position is, for example, a cross-correlation method.

Let M (M being an integer equal to or greater than 2) represent the number of microphones and τ_{ij} represent a measured value of the delay time difference between signals acquired by i-th and j-th microphones 11_i and 11_j. The measured value of the delay time difference between the acquired signals

can be obtained by calculating the cross-correlation between the acquired signals and detecting its maximum peak position. Next, let the sound acquisition position of an m-th (where m=1, ..., M) microphone be represented by (x_m, y_m, z_m) and an estimated sound source position by $(\hat{X}, \hat{Y}, \hat{Z})$. A measured value $\hat{\tau}_{ij}$ of the delay time difference between the acquired signals, which is available from these positions, is expressed by Eq. (2).

$$\hat{\tau}_{ij} = \frac{1}{c}\sqrt{(x_i - \hat{X})^2 + (y_i - \hat{Y})^2 + (z_i - \hat{Z})^2} - \frac{1}{c}\sqrt{(x_j - \hat{X})^2 + (y_j - \hat{Y})^2 + (z_j - \hat{Z})^2} \quad (2)$$

where c is sound velocity.

Next, measured and estimated value τ_{ij} and $\hat{\tau}_{ij}$ of the delay time difference between the acquired signals are multiplied by the sound velocity c for conversion into distance values, which are used as measured and estimated values d_{ij} and $\hat{d}_{ij}$ of the difference in the distance to the uttering sound source between the positions of microphones acquiring the speech sound therefrom, respectively; a mean square error e(**q**) of these values is given by Eq. (3).

$$\begin{aligned} e(\mathbf{q}) &= \sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\left|d_{ij} - \hat{d}_{ij}\right|^2 \\ &= \sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\left|d_{ij} - \sqrt{(x_i - \hat{X})^2 + (y_i - \hat{Y})^2 + (z_i - \hat{Z})^2} + \sqrt{(x_j - \hat{X})^2 + (y_j - \hat{Y})^2 + (z_j - \hat{Z})^2}\right|^2 \\ &= \sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\left|d_{ij} - r_i + r_j\right|^2 \end{aligned} \quad (3)$$

where $\mathbf{q} = (\hat{X}, \hat{Y}, \hat{Z})$. r_i and r_j represent the distances between the estimated sound source position $\mathbf{q} = (\hat{X}, \hat{Y}, \hat{Z})$ and the microphones 11_i and 11_j.

By obtaining a solution that minimizes the mean square error e(**q**) of

Eq. (3), it is possible to obtain estimated sound source position that minimizes the error between the measured and estimated values of the delay time difference between the acquired signals. In this instance, however, since Eq. (3) is nonlinear simultaneous equations and is difficult to solve analytically, the estimated sound source position is obtained by a numerical analysis using successive correction.

To obtain the estimated sound source position $(\hat{X}, \hat{Y}, \hat{Z})$ that minimizes Eq. (3), the gradient at a certain point in Eq. (3) is calculated, then the estimated sound source position is corrected in the direction in which to reduce the error until the gradient becomes zero; accordingly, the estimated sound source position is corrected by repeatedly calculating the following equation (4) for u=0, 1, ...

$$\mathbf{q}_{(u+1)} = \mathbf{q}_{(u)} - \alpha \cdot \text{grad}\, e(\mathbf{q})\big|_{q=q_{(u)}} \tag{4}$$

where α is a step size of correction, and it is set at a value $\alpha>0$. $\mathbf{q}_{(u)}$ represents $\mathbf{q}$ corrected u times, and $\mathbf{q}_{(0)}=(\hat{X}_0, \hat{Y}_0, \hat{Z}_0)$ is a predetermined arbitrary initial value when u=0. grad represents the gradient, which is expressed by the following equations (5) to (10).

$$\text{grad}\, e(\mathbf{q}) = \left(\frac{\partial e(\mathbf{q})}{\partial \hat{X}}, \frac{\partial e(\mathbf{q})}{\partial \hat{Y}}, \frac{\partial e(\mathbf{q})}{\partial \hat{Z}} \right) \tag{5}$$

$$\frac{\partial e(\mathbf{q})}{\partial \hat{X}} = 2 \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \{d_{ij} - r_i + r_j\} \cdot \left\{ \frac{x_i - \hat{X}}{r_i} - \frac{x_j - \hat{X}}{r_j} \right\} \tag{6}$$

$$\frac{\partial e(\mathbf{q})}{\partial \hat{Y}} = 2 \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \{d_{ij} - r_i + r_j\} \cdot \left\{ \frac{y_i - \hat{Y}}{r_i} - \frac{y_j - \hat{Y}}{r_j} \right\} \tag{7}$$

$$\frac{\partial e(\mathbf{q})}{\partial \hat{Z}} = 2 \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \{d_{ij} - r_i + r_j\} \cdot \left\{ \frac{z_i - \hat{Z}}{r_i} - \frac{z_j - \hat{Z}}{r_j} \right\} \tag{8}$$

$$r_i = \sqrt{(x_i - \hat{X})^2 + (y_i - \hat{Y})^2 + (z_i - \hat{Z})^2} \tag{9}$$

$$r_j = \sqrt{(x_j - \hat{X})^2 + (y_j - \hat{Y})^2 + (z_j - \hat{Z})^2} \qquad (10)$$

As described above, by repeatedly calculating Eq. (4), it is possible to obtain the estimated sound source position $\mathbf{q}=(\hat{x}, \hat{y}, \hat{z})$ where the error is minimized.

Fig. 3 illustrates in block form the functional configuration of the sound source position detecting part 15. In this example, the sound source position detecting part 15 comprises a delay time difference measuring part 15A, a multiplier 15B, a distance calculating part 15C, a mean square error calculating part 15D, a gradient calculating part 15E, a relative decision part 15F, and an estimated position updating part 15G.

The delay time difference measuring part 15A measures, during utterance from one speech sound source 9_k, the delay time difference τ_{ij} by the cross-correlation scheme for every pair (i, j) of

$i = 1, 2, \ldots, M-1;$

$j = i+1, i+2, \ldots, M$

based on the received signals by the microphones 11_i and 11_j. The multiplier 15B multiplies each measured delay time difference τ_{ij} by the sound velocity c to obtain the difference in distance, d_{ij}, between the sound source and the microphones 11_i and 11_j. The distance calculating part 15C calculates, by Eqs. (9) and (10), the distances r_i and r_j between the estimated sound source position $(\hat{x}, \hat{y}, \hat{z})$ fed from the estimated position updating part 15G and the microphones 11_i and 11_j. In this case, however, the estimated position updating part 15G provides an arbitrary initial value $(\hat{X}_0, \hat{Y}_0, \hat{Z}_0)$ as a first estimated sound source position to the distance calculating part 15C. The mean square error calculating part 15D uses d_{ij}, r_i and r_j to calculate the mean square error $e(\mathbf{q})$ by Eq. (3) for all of the above-mentioned pairs (i, j). The

gradient calculating part 15F uses the current estimate sound source position and d_{ij}, r_i, r_j to calculate the gradient grad e(**q**) of the mean square error e(**q**) by Eqs. (6), (7) and (8).

The relative decision part 15F compares each element of the gradient grad e(**q**) of the mean square error with a predetermined threshold value e_{th} to decide whether every element is smaller than the threshold value e_{th}, and if so, then outputs the estimated sound source position $(\hat{X}, \hat{Y}, \hat{Z})$ at that time. If every element is not smaller than e_{th}, then the estimated position updating part 15G uses the gradient grad e(**q**) and the current estimated position **q** = $(\hat{X}, \hat{Y}, \hat{Z})$ to update the estimated position by Eq. (4), and provides th updated estimated position $\mathbf{q}_{u+1} = (\hat{X}, \hat{Y}, \hat{Z})$ to the distance calculating part 15C. The distance calculating part 15C uses the updated estimated position $(\hat{X}, \hat{Y}, \hat{Z})$ and d_{ij} to calculate r_i and r_j updated in the same manner as referred to previously; thereafter, the mean square error calculating part 15D updates e(**q**), then the gradient calculating part 15E calculates the updated grad e(**q**), and the relative decision part 15F decides whether the updated mean square error e(**q**) is smaller than the threshold value e_{th}.

In this way, updating of the estimated position $(\hat{X}, \hat{Y}, \hat{Z})$ is repeated until every element of the gradient grad e(**q**) of the mean square error becomes sufficiently small (smaller than e_{th}), thereby estimating the position $(\hat{X}, \hat{Y}, \hat{Z})$ of the sound source 9_k. Similarly, positions of other sound sources are estimated.

The frequency domain converting part 16 converts the signal acquired by each microphone to a frequency domain signal. For example, the sampling frequency of the acquired signal is 16 kHz, and acquired signal samples from each microphone 11_m (m=1, ..., M) are subjected to FFT (Fast Fourier Transform) processing every frame of 256 samples to obtain the same

number of frequency domain signal samples $X_m(\omega)$.

Next, the covariance matrix calculating part 17 calculates the covariance of the microphone acquired signals and generates a covariance matrix. Letting $X_1(\omega)$ to $X_M(\omega)$ represent the frequency domain converted signals of the microphone acquired signals obtained by the frequency domain converting part 16 for each sound source 9_k, an M×M covariance matrix $R_{XX}(\omega)$ of these signals is generally expressed by the following equation (11).

$$\mathbf{R}_{XX}(\omega)=\begin{pmatrix} X_1(\omega) \\ \vdots \\ X_M(\omega) \end{pmatrix}\begin{pmatrix} X_1(\omega)^* & \cdots & X_M(\omega)^* \end{pmatrix}$$

$$=\begin{pmatrix} X_1(\omega)X_1(\omega)^* & X_1(\omega)X_2(\omega)^* & \cdots & X_1(\omega)X_M(\omega)^* \\ X_2(\omega)X_1(\omega)^* & X_2(\omega)X_2(\omega)^* & \cdots & X_2(\omega)X_M(\omega)^* \\ \vdots & \vdots & \ddots & \vdots \\ X_M(\omega)X_1(\omega)^* & X_M(\omega)X_2(\omega)^* & \cdots & X_M(\omega)X_M(\omega)^* \end{pmatrix} \quad (11)$$

where * represents a complex conjugate.

Next, the covariance matrix storage part 18 stores, based on the result of detection by the sound source position detecting part 15, the covariance matrix $R_{XX}(\omega)$ as an M×M covariance matrix $\mathbf{R}_{SkSk}(\omega)$ for each sound source 9_k.

Letting $\mathbf{A}_k(\omega)=(a_{k1}(\omega), \ldots, a_{kM}(\omega))$ represent the weighted mixing vectors for M-channel acquired signals for each sound source 9_k, the acquired sound level estimating part 19 calculates the acquired sound level P_{Sk} for each sound source by the following equation (12) using the covariance matrix $\mathbf{R}_{SkSk}(\omega)$ of the acquired signal for each sound source 9_k stored in the covariance matrix storage part 18.

$$P_{Sk}=\frac{1}{W}\sum_{\omega=0}^{W}\mathbf{A}_k(\omega)^H\mathbf{R}_{SkSk}(\omega)\mathbf{A}_k(\omega) \quad (12)$$

In the above, the weighted mixing vector is expressed as a vector $\mathbf{A}_k(\omega)=(a_{k1}(\omega), \ldots, a_{kM}(\omega))$ that has a controllable frequency characteristics, but if no frequency characteristics control is effected, the elements of the vector $\mathbf{A}_k$ may be preset values $a_{k1}, a_{k2}, \ldots, a_{kM}$. For example, the elements of the weighted mixing vector $\mathbf{A}_k$ for each sound source 9_k are given greater values as the microphones corresponding to the elements become closer to the sound source 9_k. In the extreme, it is possible to set 1 for only the element corresponding to the microphone 11_m closest to the sound source 9_k and set 0 for all the other elements like $\mathbf{A}_k=(0, \ldots 0, a_{km}=1,0,\ldots,0)$. In the following description, $a_{k1}(\omega), \ldots, a_{kM}(\omega)$ will be expressed simply as $a_{k1}, \ldots,, a_{kM}$, for the sake of brevity.

^{H} in Eq. (12) represents a complex conjugate transpose, and $\mathbf{A}_k(\omega)^H\mathbf{R}_{SkSk}(\omega)\mathbf{A}_k(\omega)$ can be expanded as given by the following equation.

$$\begin{aligned}
&\mathbf{A}_k(\omega)^H\mathbf{R}_{SkSk}(\omega)\mathbf{A}_k(\omega)\\
&= a_{k1}*(a_{k1}X_1(\omega)X_1(\omega)*+ a_{k2}X_2(\omega)X_1(\omega)*+\ldots+ a_{kM}X_M(\omega)X_1(\omega)*)\\
&\quad + a_{k2}*(a_{k1}X_1(\omega)X_2(\omega)*+ a_{k2}X_2(\omega)X_2(\omega)*+\ldots+ a_{kM}X_M(\omega)X_2(\omega)*)\\
&\qquad\qquad\qquad\qquad\vdots\qquad\qquad\qquad\qquad\vdots\\
&\quad + a_{kM}*(a_{k1}X_1(\omega)X_M(\omega)*+ a_{k2}X_2(\omega)X_M(\omega)*+\ldots+a_{kM}X_M(\omega)X_M(\omega)*)\\
&=\Omega(\omega) \qquad\qquad (13)
\end{aligned}$$

Eq. (12) means that the mean power P_{sk} of the acquired signal is calculated by adding the power spectrum sample value represented by $\Omega(\omega)$ given by Eq. (13) for bands 0 to W (sample number) of the frequency domain signal generated by the frequency domain converting part 16 and then dividing the added value by W.

For example, assuming that the microphone 11_1 is the closest to the sound source 9_1, the value of the weighting factor a_{k1} is so determined as to assign the maximum weight to the acquired signal by the microphone 11_1 (a

first channel) and the values of weighting factors a_{k2}, a_{k3}, ..., a_{kM} for the acquired signals of the other channels are determined smaller than a_{k1}. With such weighting scheme, it is possible to increase S/N of the acquired signal from the sound source 9_1 or lessen the influence of room reverberation more than in the case where such weighting is not performed. That is, the optimum value of the weighting factor of the weighted mixing vector $\mathbf{A}_k(\omega)$ for each sound source 9_k is predetermined experimentally by the directivity and layout of microphones and the layout of sound sources in such a manner as to increase S/N of the output speech signal corresponding to the sound source 9_k, for example, and decrease the room reverberation. According to the present invention, however, even when equal weighting is done in all the channels, acquired signals from the respective sound sources can be controlled to a desired level.

Next, the filter coefficient calculating part 21 calculates filter coefficients for acquiring speech sound from each sound source in a desired volume. In the first place, let $H_1(\omega)$ to $H_M(\omega)$ represent frequency domain converted versions of filter coefficients of the filters 12_1 to 12_M each connected to one of the microphones. Next, let $\mathbf{H}(\omega)$ represent a matrix of these filter coefficients by the following equation (14).

$$\mathbf{H}(\omega) = \begin{pmatrix} H_1(\omega) \\ \vdots \\ H_M(\omega) \end{pmatrix} \qquad (14)$$

Further, let $X_{Sk,1}$ to $X_{Sk,M}$ represent frequency domain converted signals of the signals acquired by respective microphones during utterance of the k-th sound source 9_k.

In this case, the condition that the filter coefficient matrix $\mathbf{H}(\omega)$ needs to satisfy is that when the microphone acquired signals are subjected to

filtering with the filter coefficient matrix $\mathbf{H}(\omega)$ and the filtered signals are added together, the signal component from each sound source has a desired level P_{opt}. Accordingly, the following equation (15) is an ideal condition by which the signal obtained by adding the filtered signals from the sound source 9_k becomes equivalent to a signal obtained by multiplying the weighted mixing vector $\mathbf{A}_k(\omega)$ for the acquired signals from the microphones 11_1 to 11_M by a desired gain.

$$\left(X_{Sk,1}(\omega)\cdots X_{Sk,M}(\omega)\right)\mathbf{H}(\omega)=\sqrt{\frac{P_{opt}}{P_{Sk}}}\left(X_{Sk,1}(\omega)\cdots X_{Sk,M}(\omega)\right)\mathbf{A}_k(\omega) \qquad (15)$$

where k=1, ..., K, K representing the number of sound sources.

Next, solving the condition of Eq. (15) by the least square method for the filter coefficient matrix $\mathbf{H}(\omega)$ gives the following equation (16).

$$\mathbf{H}(\omega)=\left\{\sum_{k=1}^{K}C_{Sk}\mathbf{R}_{SkSk}(\omega)\right\}^{-1}\sum_{k=1}^{K}C_{Sk}\sqrt{\frac{P_{opt}}{P_{Sk}}}\mathbf{R}_{SkSk}(\omega)\mathbf{A}_k(\omega) \qquad (16)$$

where C_{Sk} is a weighting factor that imposes a sensitivity restraint on the k-th sound source position. The sensitivity restraint mentioned herein means flattening the frequency characteristics of the present sound acquisition apparatus with respect to the sound source position. An increase in this value increases the sensitivity restraint on the sound source concerned, permitting sound acquisition with flatter frequency characteristics but increasing deterioration of frequency characteristics for other sound source positions. Hence, it is preferable that C_{Sk} be normally set at a value approximately in the range of 0.1 to 10 to impose well-balances restraints on all the sound sources.

Fig. 4 illustrates in block form the functional configuration of the filter coefficient calculating part 21 for calculating the filter coefficients expressed by Eq. (16). In this example, the covariance matrices $\mathbf{R}_{S1S1}$ to $\mathbf{R}_{SKSK}$

corresponding to the respective sound sources 9_1 to 9_K, provided from the covariance matrix storage part 18, are applied to multipliers 21A1 to 21Ak, wherein they are multiplied by weighting factors C_{S1} to C_{SK} set by a weighting factor setting part 21H, respectively. The acquired sound levels P_{S1} to P_{SK} for the sound sources 9_1 to 9_K, estimated by the acquired sound level estimating part 19, are provided to square ratio calculating parts 21B1 to 21BK, wherein square ratios, $(P_{opt}/P_{S1})^{1/2}$ to $(P_{opt}/P_{SK})^{1/2}$, between them and the predetermined desired output level P_{opt} are calculated, and the calculated values are provided to multipliers 21C1 to 21CK for multiplication by the outputs from multipliers 21A1 to 21AK, respectively. The outputs from the multipliers 21C1 to 21CK are fed to multipliers 21D1 to 21DK, where they are further multiplied by weighted mixing vectors $\mathbf{A}_1(\omega)$ to $\mathbf{A}_K(\omega)$, and a matrix of the total sum of the multiplied outputs is calculated by an adder 21E. On the other hand, a matrix of the sum total of the outputs from the multipliers 21A1 to 21AK is calculated by an adder 21F, and by an inverse matrix multiplier 21G, an inverse matrix of the matrix calculated by the adder 21F and the output from the adder 21E are multiplied to calculate the filter coefficient $\mathbf{H}(\omega)$.

Next, the filter coefficients $H_1(\omega)$, $H_2(\omega)$, ..., $H_M(\omega)$ calculated by the filter coefficient calculating part 21 are set in the filters 12_1 to 12_M for filtering the acquired signals from the microphones 11_1 to 11_M, respectively. The filtered signals are added together by the adder 13, from which the added output is provided as an output signal.

A description will be given below of three examples of the usage of the sound acquisition apparatus according to the present invention.

A first method begins, as shown in Fig. 5, with initial setting of the number K of sound sources at K=0 in step S1. This is followed by step S2

in which the state decision part 14 periodically makes a check for utterance, and if utterance is detected, the sound source position detecting part 15 detects the position of the sound source concerned in step S3. In step S4 it is decided whether the detected sound source position matches any one of those previously detected, and if a match is found, the covariance matrix $\mathbf{R}_{XX}(\omega)$ corresponding to that sound source position is newly calculated in the covariance matrix calculating part 17 in step S5, and in step S6 the covariance matrix in the corresponding area of the covariance matrix storage part 18 is updated with the newly calculated covariance matrix.

When no match is found with the previously detected sound source position in step S4, K is incremented by 1 in step S7, then in step S8 the covariance matrix $\mathbf{R}_{XX}(\omega)$ corresponding to that sound source position is newly calculated in the covariance matrix calculating part 17, and in step S9 the covariance matrix is stored in a new area of the covariance matrix storage part 18.

Next, in step S10 the acquired sound level is estimated from the stored covariance matrix in the acquired sound level estimating part 19, then in step S11 the estimated acquired sound level and the covariance matrix are used to calculate the filter coefficients $H_1(\omega)$ to $H_M(\omega)$ by the filter coefficient calculating part 17, and in step S12 the filter coefficients set in the filters 12_1 to 12_M are updated with the newly calculated ones.

A second method begins, as shown in Fig. 6, with presetting the maximum value of the sound source number at K_{max} and presetting the initial value of the sound source number K at 0 in step S1. The subsequent steps S2 to S6 are the same as in the case with Fig. 5; that is, the microphone output signals are checked for utterance, and if utterance is detected, then its sound source position is detected, then it is decided whether the detected sound

source position matches any one of those previously detected, and if a match is found, the covariance matrix corresponding to that sound source position is calculated and stored as a newly updated matrix in the corresponding storage area.

When it is found in step S4 that the detected sound source position does not match any one of previously detected positions, K is incremented by 1 in step S7, and in step S8 a check is made to see if K is larger than the maximum value K_{max}. If it does not exceed the maximum value K_{max}, then the covariance matrix for the detected position is calculated in step S9, and in step S10 the covariance matrix is stored in a new area. When it is found that in step S8 that K exceeds the maximum value K_{max}, $K=K_{max}$ is set in step S11, then in step S12 the most previously updated one of the covariance matrices stored in the covariance matrix storage part 18 is erased, and a new covariance matrix calculated by the covariance matrix calculating part 17 in step S13 is stored in that area in step S14. The subsequent steps S15, S16 and S17 are the same as in steps S10, S11 and S12 in Fig. 5; that is, the estimated acquired sound level for each sound source is calculated from the covariance matrix, and filter coefficients are calculated and set in the filters 12_1 to 12_M. This method is advantageous over the Fig. 5 method in that the storage area of the covariance matrix storage part 18 can be reduced by limiting the maximum value of the sound source number K to K_{max}.

In the first and second methods, described above, each detection of speech sound is always accompanied by the calculation and storage of a covariance matrix and updating of the filter coefficients, but the third method described below does not involve updating of the filter coefficients when the position of the sound source of the detected utterance matches any one of the already detected sound source positions. Fig. 7 illustrates the procedure of

the third method. In step S1 the initial value of the sound source number k is set to 0, then in step S2 the state detecting part 14 periodically makes a check for utterance, and if utterance is detected, the sound source position detecting part 15 detects the position of the sound source of the detected utterance in step S3. In step S4 it is decided whether the detected sound source position matches any one of the already detected sound source positions, and if a match is found, the procedure returns to step S2 without updating. If no match is found with any one of the already detected sound source positions in step S4, that is, if the sound source 9_k moves to a position different from that where it was previously, or if a new sound source is added, K is incremented by 1 in step S5, then in step S6 the covariance matrix $\mathbf{R}_{SkSk}(\omega)$ corresponding to the sound source is newly calculated in the covariance calculating part 17, and in step S7 it is stored in the corresponding new area MA_k of the covariance storage part 18, then in step S8 the covariance matrix is used to estimate the acquired sound level by the acquired sound level estimating part 19, then in step S9 all the covariance matrices and estimated acquired sound levels to calculate updated filter coefficients by the filter coefficient calculating part 21, and in step S10 the updated filter coefficients are set in the filters 12_1 to 12_M, followed by a return to step S2.

As described above, according to the present invention, the sound source positions are estimated from the acquired signals of a plurality of microphones, then a covariance matrix of the acquired signal is calculated for each sound source, then filter coefficients for adjusting the sound volume for each sound source position are calculated, and the filter coefficients are used to filter the acquired signals of the microphones, by which it is possible to obtain an output signal of a volume adjusted for each speaker's position.

While the Fig. 1 embodiment has been described with reference to the

case where the sound source position detecting part 15 estimates the coordinate position of each sound source 9_k, it is also possible to calculate the sound source direction, that is, the angular position of each sound source 9_k to the arrangement of the microphones 11_1 to 11_M. A method for estimating the sound source direction is set forth, for example, in Tanaka, Kaneda, and Kojima, "Performance Evaluation of a Sound Source Direction Estimating Method under Room Reverberation," Journal of the Society of Acoustic Engineers of Japan, Vol. 50, No. 7, 1994, pp. 540-548. In short, a covariance matrix of acquired signals needs only to be calculated for each sound source and stored.

SECOND EMBODIMENT

Fig. 8 is a functional block diagram of a sound acquisition apparatus according to a second embodiment of the present invention.

The sound acquisition apparatus of this embodiment comprises microphones 11_1 to 11_M, filters 12_1 to 12_M, an adder 13, a state decision part 14, a sound source position detecting part 15, a frequency domain converting part 16, a covariance matrix calculating part 17, a covariance matrix storage part 18, an acquired sound level estimating part 19, and a filter coefficient calculating part 21.

This embodiment adds an effect of noise reduction to the acquired sound level adjustment of the sound acquisition apparatus according to the first embodiment of the present invention.

In the first place, the state decision part 14 detects an utterance period and a noise period from the power of the received signals from the microphones 11_1 to 11_M. The state decision part 14 includes, as shown in Fig. 9, a noise decision part 14F added to the state decision part 14 of Fig. 2. For example, as is the case with the first embodiment, a short-time mean

power P_{avS} and a long-time mean power P_{avL} are calculated by the short-time mean power calculating part 14B and the long-time mean power calculating part 14C for the acquired signals from the respective microphones, then the ratio, $R_p=P_{avS}/P_{avL}$, between the short-time mean power and the long-time mean power is calculated in the division part 14D, then the ratio is compared with an utterance threshold value P_{thU} in the utterance decision part 14E, and if the power ratio exceeds the threshold value, it is decided as indicating the presence of the utterance period. The noise decision part 14F compares the power ratio R_p with a noise threshold value P_{thN}, and if the power ratio is smaller than the threshold value, it is decided as indicating the presence of the noise period.

When the result of decision by the utterance decision part 14E is indicative of the utterance period, the sound source position detecting part 15 detects the position of the sound source concerned in the same way as in the first embodiment of the present invention.

Next, the frequency domain converting part 16 converts acquired signals from the microphones 11_1 o 11_M in the utterance period and in the noise period of each sound source 9_k into frequency domain signals, and provides them to the covariance matrix calculating part 17. The covariance matrix calculating part 17 calculates a covariance matrix $\mathbf{R}_{SkSk}(\omega)$ of the frequency domain acquired signals for the sound source 9_k in the same manner as in the first embodiment of the present invention. Further, the covariance matrix calculating part calculates a covariance matrix $\mathbf{R}_{NN}(\omega)$ of the frequency domain acquired signals in the noise period.

The covariance matrix storage part 18 stores, based on the result of detection by the sound source position detecting part 15 and the result of decision by the state decision part 15, the covariance matrices $\mathbf{R}_{SkSk}(\omega)$ in the

utterance period and the covariance matrices $\mathbf{R}_{NN}(\omega)$ in the noise period for the sound sources $9_1, \ldots, 9_K$ in areas $MA_1, \ldots, MA_K, MA_{K+1}$.

The acquired sound level estimating part 19 estimates the acquired sound level P_{Sk} for each sound source in the same manner as in the first embodiment of the present invention.

Next, the filter coefficient calculating part 21 calculates filter coefficients for acquiring sound from each sound source 9_k at a desired volume and for reducing noise. In the first place, the condition for noise reduction is calculated. Let the frequency domain converted signals of the microphone acquired signals in the noise period be represented by $X_{N,1}(\omega)$ to $X_{N,M}(\omega)$. If the microphone acquired signals $X_{N,1}(\omega)$ to $X_{N,M}(\omega)$ in the noise period become zero after passing through the filters 12_1 to 12_M and the adder 13, this means that noise could be reduced; hence, the condition for noise reduction is given by the following equation (17).

$$(X_{N,1}(\omega), \ldots, X_{N,M}(\omega))\mathbf{H}(\omega) = 0 \qquad (17)$$

By satisfying both of Eq. (17) and Eq. (15) for adjusting the acquired sound level, mentioned previously in the first embodiment of the present invention, it is possible to implement both of the acquired sound level adjustment and the noise reduction.

Next, solving the conditions of Eqs. (15) and (17) by the least square method for the filter coefficient matrix $\mathbf{H}(\omega)$ gives the following equation (18).

$$\mathbf{H}(\omega) = \left\{ \sum_{k=1}^{K} C_{Sk} \mathbf{R}_{SkSk}(\omega) + C_N \mathbf{R}_{NN}(\omega) \right\}^{-1} \sum_{k=1}^{K} C_{Sk} \sqrt{\frac{P_{opt}}{P_{Sk}}} \mathbf{R}_{SkSk}(\omega) \mathbf{A}_k(\omega) \qquad (18)$$

C_N is a weight constant for the noise reduction rate; an increase in the value of the constant increases the noise reduction rate. But, since an increase in C_N

decreases the sensitivity constraint for the sound source position and increases degradation of the frequency characteristics of the acquired sound signal, CN is normally set at an appropriate value approximately in the range of 0. 1 to 10.0. The meanings of the other symbols are the same as in the first embodiment.

Next, the filter coefficients calculated by Eq. (18) are set in the filters 12_1 to 12_M and used to filter the microphone acquired signals. The filtered signals are added together by the adder 13, from the added signal is provided as the output signal.

As described above, the second embodiment of the present invention permits reduction of noise in addition to the effect of acquired sound level adjustment in the first embodiment of the present invention.

The other parts of this embodiment are the same as in the first embodiment of the present invention, and hence they will not be described.

THIRD EMBODIMENT

Fig. 10 is a functional block diagram of a sound acquisition apparatus according to a third embodiment of the present invention.

The sound acquisition apparatus of this embodiment comprises a loudspeaker 22, microphones 11_1 to 11_M, filters 12_1 to 12_M and 23, an adder 13, a state decision part 14, a sound source position detecting part 15, a frequency domain converting part 16, a covariance matrix calculating part 17, a covariance matrix storage part 18, an acquired sound level estimating part 19, and a filter coefficient calculating part 21.

This embodiment adds, to the sound acquisition apparatus of the second embodiment of the present invention, the loudspeaker 22 for reproducing a received signal from a participating speaker at a remote location and the filter 23 for filtering the received signal, with a view to

implementing, in addition to the acquired sound level adjustment and the noise reduction by the second embodiment, cancellation of acoustic echoes that are loudspeaker reproduced signal components which are acquired by the microphones 11_1 to 11_M.

The state decision part 14 has, in addition to the Fig. 4 configuration of the state decision part 14 as shown in Fig. 11: a short-time mean power calculating part 14B' and a long-time mean power calculating part 14C' for calculating short-time mean power P'_{avS} and long-time mean power P'_{avL} of the received signal, respectively; a division part 14D' for calculating their ratio $R'_p=P'_{avS}/P'_{acL}$; a receive decision part 14G that compares the ratio R'_p with a predetermined receive signal threshold value R_{thR} and, if the former is larger than the latter, decides the state as a receiving period; and a state determining part 14H that determines the state based on the results of decision by the utterance decision part 14E, the noise decision part 14F and the receive decision part 14G. When the result of decision by the receive decision part 14G is the receiving period, the state determining part 14H determines the state as the receiving period, irrespective of the results of decision by the utterance decision part 14E and the noise decision part 14F, whereas when the receive decision part 14G decides that the state is not the receiving period, the state determining part determines the state as the utterance or noise period according to the decisions by the utterance decision part 14E and the noise decision part 14F as in the case of Fig. 4.

When the result of decision by the state decision part 14 is the utterance period, the sound source position detecting part 15 detects the position of the sound source concerned in the same manner as in the first embodiment of the present invention.

Next, the frequency domain converting part 16 converts the

microphone acquired signals and the received signal to frequency domain signals $X_1(\omega)$, ..., $X_M(\omega)$ and $Z(\omega)$, and the covariance matrix calculating part 17 calculates covariance matrices of the frequency domain acquired signals and received signal. A covariance matrix $\mathbf{R}_{XX}(\omega)$ of the frequency domain converted signals $X_1(\omega)$ to $X_M(\omega)$ of the microphone acquired signals and the frequency domain converted signal $Z(\omega)$ is calculated by the following equation (19).

$$R_{XX}(\omega) = \begin{pmatrix} Z(\omega) \\ X_1(\omega) \\ \vdots \\ X_M(\omega) \end{pmatrix} \left(Z(\omega)^* X_1(\omega)^* \cdots X_M(\omega)^*\right) \qquad (19)$$

where * represents a complex conjugate.

Next, in the covariance matrix storage part 18, based on the result of detection by the sound source position detecting part 15 and on the result of decision by the state decision part 14, the covariance matrix $\mathbf{R}_{XX}(\omega)$ is stored as covariance matrices $\mathbf{R}_{SkSk}(\omega)$ of the acquired signals and the received signal for each sound source 9_k in the utterance period, as a covariance matrix $\mathbf{R}_{NN}(\omega)$ of the acquired signals and the received signal in the noise period, and as a covariance matrix $\mathbf{R}_{EE}(\omega)$ of the acquired signals and the received signal in the receiving period in areas MA_1, ..., MA_K, MA_{K+1}, MA_{K+2}, respectively.

The acquired sound level estimating part 19 calculates the acquired sound level P_{Sk} for each sound source 9_k by the following equation (20) based on the covariance matrices $\mathbf{R}_{S1S1}$, ..., $\mathbf{R}_{SKSK}$ for each sound source and predetermined weighted mixing vectors $\mathbf{A}_1(\omega)$, ..., $\mathbf{A}_K(\omega)$ composed of M+1 elements for each sound source.

$$P_{Sk} = \frac{1}{W} \sum_{\omega=0}^{W} \mathbf{A}_k(\omega)^H R_{SkSk}(\omega) \mathbf{A}_k(\omega) \qquad (20)$$

Next, the filter coefficient calculating part 21 calculates filter coefficients for acquiring, at a desired volume, speech sound uttered from each sound source. Let $\mathbf{H}_1(\omega)$ to $\mathbf{H}_M(\omega)$ represent frequency domain converted versions of the filter coefficients of the filters 12_1 to 12_M connected to the microphones, respectively, and let $F(\omega)$ represent a frequency domain converted version of the filter coefficient of the filter 23 for filtering the received signal. Then, let $\mathbf{H}(\omega)$ represent a matrix of these filter coefficients given by the following equation (21).

$$\mathbf{H}(\omega) = \begin{pmatrix} F(\omega) \\ H_1(\omega) \\ \vdots \\ H_M(\omega) \end{pmatrix} \tag{21}$$

Further, let $X_{E,1}(\omega)$ to $X_{E,M}(\omega)$ represent frequency domain converted signals of the microphone acquired signal in the receiving period; let $Z_E(\omega)$ represent a frequency domain converted signal of the received signal; let $X_{N,1}(\omega)$ to $X_{N,M}(\omega)$ represent frequency domain converted signals of the microphone acquired signals in the noise period; let $Z_N(\omega)$ represent a frequency domain converted signal of the received signal; let $X_{Sk,1}(\omega)$ to $X_{Sk,M}(\omega)$ represent frequency domain converted signals of the microphone acquired signals in the utterance period of the k-th sound source 9_k; and let $Z_{Sk}(\omega)$ represent a frequency domain converted signal of the received signal.

In this case, the condition that the filter coefficient matrix $\mathbf{H}(\omega)$ needs to meet is that when the microphone acquired signals and the send signal are each subjected to filtering with the filter coefficient matrix $\mathbf{H}(\omega)$ and the signals after filtering are added together, acoustic echo and noise signals are cancelled and only the send speech signal is sent at a desired level.

Accordingly, for the signals during the receiving period and the noise

period, the following equations (22) and (23) are ideal conditions by which the filtered and added signals become 0.

$$\begin{pmatrix} Z_E(\omega) & X_{E,1}(\omega) & \cdots & X_{E,M}(\omega) \end{pmatrix} \mathbf{H}(\omega) = 0 \tag{22}$$

$$\begin{pmatrix} Z_N(\omega) & X_{N,1}(\omega) & \cdots & X_{N,M}(\omega) \end{pmatrix} \mathbf{H}(\omega) = 0 \tag{23}$$

For the signal during the utterance period, the following equation is an ideal condition by which the filtered and added signal becomes equivalent to a signal obtained by multiplying the microphone acquired signals and the received signal by the weighted mixing vector $\mathbf{A}_k(\omega)$ composed of predetermined M+1 elements and a desired gain.

$$\begin{pmatrix} Z_{Sk}(\omega) & X_{Sk,1}(\omega) & \cdots & X_{Sk,M}(\omega) \end{pmatrix} \mathbf{H}(\omega) = \sqrt{\frac{P_{opt}}{P_{Sk}}} \begin{pmatrix} Z_{Sk}(\omega) & X_{Sk,1}(\omega) & \cdots & X_{Sk,M}(\omega) \end{pmatrix} \mathbf{A}_k(\omega) \tag{24}$$

The element $a_0(\omega)$ of the weighted mixing vector $\mathbf{A}_k(\omega)=(a_0(\omega),a_{k1}(\omega), \ldots, a_{kM}(\omega))$ represents a weighting factor for the received signal; normally, it is set at $a_0(\omega)=0$.

Next, solving the conditions of Eqs. (22) to (24) by the least square method for the filter coefficient matrix $\mathbf{H}(\omega)$ gives the following equation:

$$\mathbf{H}(\omega) = \left\{ \sum_{k=1}^{K} C_{Sk} \mathbf{R}_{SkSk}(\omega) + C_N \mathbf{R}_{NN}(\omega) + C_E \mathbf{R}_{EE}(\omega) \right\}^{-1} \sum_{k=1}^{K} C_{Sk} \sqrt{\frac{P_{opt}}{P_{Sk}}} \mathbf{R}_{SkSk}(\omega) \mathbf{A}_k(\omega) \tag{25}$$

C_E is a weight constant for acoustic echo return loss enhancement; the larger the value, the more the acoustic echo return loss enhancement increases. But, an increase in the value C_E accelerates deterioration of the frequency characteristics of the acquired signal and lowers the noise reduction characteristic. On this account, C_E is usually set at an appropriate value approximately in the range of 0.1 to 10.0. The meanings of the other symbols are the same as in the second embodiment.

In this way, the filter coefficients can be determined in such a manner as to adjust volume and reduce noise.

Next, the filter coefficients, obtained by Eq. (25), are set in the filters 12_1 to 12_M and 23, which filter the microphone acquired signals and the received signal, respectively. The filtered signals are added together by the adder 13, from which the added signal is output as the send signal. The other parts are the same as in the second embodiment of the present invention, and hence no description will be repeated.

As described above, the third embodiment of the present invention permits implementation of acoustic echo cancellation in addition to the effects of acquired sound level adjustment and noise reduction by the second embodiment of the present invention. While the third embodiment has been described as adding the acoustic echo cancellation capability to the second embodiment, the acoustic echo cancellation capability may also be added to the first embodiment. In such an instance, the noise decision part 14F is removed in Fig. 11 showing in detail the state decision part 14 in Fig. 10, and the covariance matrix calculating part 17 in Fig. 10 does not calculate the covariance matrix $\mathbf{R}_{NN}(\omega)$ in the noise period. Accordingly, the calculation of filter coefficients in the filter coefficient calculating part 21 may be carried out by the following equation, which is evident from the foregoing description.

$$\mathbf{H}(\omega)=\left\{\sum_{k=1}^{K} C_{Sk}\mathbf{R}_{SkSk}(\omega)+C_E\mathbf{R}_{EE}(\omega)\right\}^{-1}\sum_{k=1}^{K} C_{Sk}\sqrt{\frac{P_{opt}}{P_{Sk}}}\mathbf{R}_{SkSk}(\omega)\mathbf{A}_k(\omega) \tag{26}$$

FOURTH EMBODIMENT

Though described above as an embodiment having added the acoustic echo cancellation capability to the acquired sound level adjustment and noise

reduction capabilities of the second embodiment, the third embodiment of Fig. 10 may also be configured as a sound acquisition apparatus equipped with only the noise reduction and acoustic echo cancellation capabilities. An example of such a configuration is shown in Fig. 12.

As illustrated in Fig. 12, this embodiment has a configuration in which the sound source position detecting part 15 and the acquired sound level estimating part 19 in the Fig. 10 configuration are removed and the covariance matrix calculating part 17 calculates a covariance matrix $\mathbf{R}_{SS}(\omega)$ of the send signal, a covariance matrix $\mathbf{R}_{EE}(\omega)$ of the received signal, and a covariance matrix $\mathbf{R}_{NN}(\omega)$ of the noise signal, which are stored in storage areas MA_S, MA_E and MA_N of the covariance storage part 18, respectively. The acoustic echo cancellation capability can be implemented using at least one microphone, but an example using M microphones is shown.

The state decision part 14 decides, as in the Fig. 10 embodiment, the utterance period, the receiving period, and the noise period from the signals acquired by the microphones 12_1 to 12_M and the received signal; the state decision part is identical in concrete configuration and in operation with the counterpart depicted in Fig. 11. The acquired signals and the received signal are converted by the frequency domain converting part 16 to frequency domain acquired signals $X_1(\omega)$ to $X_M(\omega)$ and a frequency domain received signal $Z(\omega)$, which are provided to the covariance matrix calculating part 17.

Next, the covariance matrix calculating part 17 generates a covariance matrix of the frequency domain acquired signals and received signal. The covariance matrix $\mathbf{R}_{XX}(\omega)$ of the frequency domain converted signals $X_1(\omega)$ to $X_M(\omega)$ of the microphone acquired signals and the frequency domain converted signal $Z(\omega)$ of the received signal is calculated by the following equation (27).

$$R_{XX}(\omega) = \begin{pmatrix} Z(\omega) \\ X_1(\omega) \\ \vdots \\ X_M(\omega) \end{pmatrix} (Z(\omega) * X_1(\omega) * \cdots X_M(\omega) *) \qquad (27)$$

where * represents a complex conjugate.

Next, in the covariance matrix storage part 18, based on the result of detection by the state decision part 14, the covariance matrix $\mathbf{R}_{XX}(\omega)$ is stored as a covariance matrix $\mathbf{R}_{SS}(\omega)$ of the acquired signals and the received signal for each sound source 9_k in the utterance period, as a covariance matrix $\mathbf{R}_{NN}(\omega)$ of the acquired signals and the received signal in the noise period, and as a covariance matrix $\mathbf{R}_{EE}(\omega)$ of the acquired signals and the received signal in the receiving period in areas MA_S, MA_N, and MA_E, respectively.

Next, the filter coefficient calculating part 21 acquires speech sound uttered from sound sources, and calculates filter coefficients for canceling acoustic echo and noise. Let $\mathbf{H}_1(\omega)$ to $\mathbf{H}_M(\omega)$ represent frequency domain converted versions of the filter coefficients of the filters 12_1 to 12_M connected to the microphones 11_1 to 11_M, respectively, and let $F(\omega)$ represent a frequency domain converted version of the filter coefficient of the filter 23 for filtering the received signal. Then, let $\mathbf{H}(\omega)$ represent a matrix of these filter coefficients given by the following equation (28).

$$\mathbf{H}(\omega) = \begin{pmatrix} F(\omega) \\ H_1(\omega) \\ \vdots \\ H_M(\omega) \end{pmatrix} \qquad (28)$$

Further, let $X_{E,1}(\omega)$ to $X_{E,M}(\omega)$ represent frequency domain converted signals of the microphone acquired signals in the receiving period; let $Z_E(\omega)$ represent a frequency domain converted signal of the received signal; let

$X_{N,1}(\omega)$ to $X_{N,M}(\omega)$ represent frequency domain converted signals of the microphone acquired signals in the noise period; let $Z_N(\omega)$ represent a frequency domain converted signal of the received signal; let $X_{Sk,1}(\omega)$ to $X_{Sk,M}(\omega)$ represent frequency domain converted signals of the microphone acquired signals in the utterance period; and let $Z_S(\omega)$ represent a frequency domain converted signal of the received signal in the utterance period.

In this case, the condition that the filter coefficient matrix $\mathbf{H}(\omega)$ needs to meet is that when the microphone acquired signals and the send signal are each subjected to filtering with the filter coefficient matrix $\mathbf{H}(\omega)$ and the signals after filtering are added together, acoustic echo and noise signals are cancelled and only the send speech signal is sent at a desired level.

Accordingly, for the signals during the received signal period and the noise period, the following equations (29) and (30) are ideal conditions by which the filtered and added signals become 0.

$$\begin{pmatrix} Z_E(\omega) & X_{E,1}(\omega) & \cdots & X_{E,M}(\omega) \end{pmatrix} \mathbf{H}(\omega) = 0 \quad (29)$$

$$\begin{pmatrix} Z_N(\omega) & X_{N,1}(\omega) & \cdots & X_{N,M}(\omega) \end{pmatrix} \mathbf{H}(\omega) = 0 \quad (30)$$

For the signal during the utterance period, the following equation is an ideal condition by which the filtered and added signal becomes equivalent to a signal obtained by multiplying the microphone acquired signals and the received signal by the weighted mixing vector $\mathbf{A}(\omega)$ composed of predetermined M+1 elements.

$$\begin{pmatrix} Z_S(\omega) & X_{S,1}(\omega) & \cdots & X_{S,M}(\omega) \end{pmatrix} \mathbf{H}(\omega) = \begin{pmatrix} Z_S(\omega) & X_{S,1}(\omega) & \cdots & X_{S,M}(\omega) \end{pmatrix} \mathbf{A}(\omega) \quad (31)$$

The first element $a_0(\omega)$ of the weighted mixing vector $\mathbf{A}(\omega)=(a_0(\omega),a_{k1}(\omega), \ldots, a_{kM}(\omega))$ represents a weighting factor for the received signal; normally, it is set at $a_0(\omega)=0$.

Next, solving the conditions of Eqs. (29) to (31) by the least square

method for the filter coefficient matrix $\mathbf{H}(\omega)$ gives the following equation:

$$\mathbf{H}(\omega) = \{\mathbf{R}_{SS}(\omega) + C_N\mathbf{R}_{NN}(\omega) + C_E\mathbf{R}_{EE}(\omega)\}^{-1}\mathbf{R}_{SS}(\omega)\mathbf{A}(\omega) \quad (32)$$

C_E is a weight constant for acoustic echo return loss enhancement; the larger the value of the weight constant, the more the acoustic echo return loss enhancement increases. But, an increase in the value C_E accelerates deterioration of the frequency characteristics of the acquired signal and lowers the noise reduction characteristic. On this account, C_E is usually set at an appropriate value approximately in the range of 0.1 to 10.0. The meanings of the other symbols are the same as in the second embodiment.

In this way, the filter coefficients can be determined in such a manner as to adjust volume and reduce noise.

Next, the filter coefficients, obtained by Eq. (32), are set in the filters 12_1 to 12_M and 23, which filter the microphone acquired signals and the received signal, respectively. The filtered signals are added together by the adder 13, from which the added signal is output as the send signal. The other parts are the same as in the second embodiment of the present invention, and hence no description will be repeated.

As described above, the fourth embodiment of the present invention permits implementation of acoustic echo cancellation in addition to the effect of noise reduction.

FIFTH EMBODIMENT

Fig. 13 illustrates a fifth embodiment. According to the fifth embodiment, in the fourth embodiment of Fig. 12, sound source positions are detected during the utterance period, a covariance matrix is calculated for each sound source and stored and during the noise period a covariance matrix for noise is calculated and stored. Then, these stored covariance matrices are

used to calculate filter coefficients for canceling noise and acoustic echo. The microphone acquired signals and the received signal are filtered using these filter coefficients to thereby obtain a send signal from which noise and acoustic echo have been cancelled.

The configuration of the fifth embodiment is common to the configuration of the third embodiment except the removal of the acquired sound level estimating part 19 in Fig. 10.

The state decision part 14 detects the utterance period, the receiving period and the noise period as in the third embodiment. When the result of decision by the state decision part 14 is the utterance period, the sound source position detecting part 15 estimates the position of each sound source 9_k. The sound source position estimating method is the same as that used in the first embodiment of Fig. 1, no description will be repeated.

Next, in the frequency domain converting part 16 the acquired signals and the received signal are converted to frequency domain signals, which are provided to the covariance matrix calculating part 17.

The covariance matrix calculating part 17 calculates covariance matrices $\mathbf{R}_{S1S1}(\omega)$ to $\mathbf{R}_{SKSK}(\omega)$ of the acquires signals for the respective sound sources 9_k and the received signal, a covariance matrix $\mathbf{R}_{EE}(\omega)$ in the receiving period and a covariance matrix $\mathbf{R}_{NN}(\omega)$ in the noise period. The covariance matrix storage part 18 stores the covariance matrices $\mathbf{R}_{S1S1}(\omega)$ to $\mathbf{R}_{SKSK}(\omega)$, $\mathbf{R}_{EE}(\omega)$ and $\mathbf{R}_{NN}(\omega)$ in the corresponding areas MA_1 to, MA_K, MA_{K+1} and MA_{K+2}, respectively, based on the result of decision by the state decision part 14 and the results of position detection by the sound source position detecting part 15.

Upon the send speech sound being acquired, the filter coefficient calculating part 21 calculates filter coefficients for canceling acoustic echo

and noise. As is the case with the third embodiment, solving the conditional expression for the filter coefficient matrix $\mathbf{H}(\omega)$ by the least square method gives the following equation:

$$\mathbf{H}(\omega) = \left\{ \sum_{k=1}^{K} C_{Sk} \mathbf{R}_{SkSk}(\omega) + C_N \mathbf{R}_{NN}(\omega) + C_E \mathbf{R}_{EE}(\omega) \right\}^{-1} \sum_{k=1}^{K} C_{Sk} \mathbf{R}_{SkSk}(\omega) \mathbf{A}_k(\omega) \quad (33)$$

In the above, C_{S1} to C_{SK} are weight constants of sensitivity constraints for the respective sound sources, C_E is a weight constant for the echo return loss enhancement, and C_N is a weight constant for the noise reduction rate.

The filter coefficients thus obtained are set in the filters 12_1 to 12_M and 23, which filter the microphone acquired sound signals and the received signal, respectively. The filtered signals are added together by the adder 13, from which the added signal is output as the send signal. The other parts are the same as in the second embodiment of the present invention, and hence no description will be repeated. The fifth embodiment permits generation of a send signal having cancelled therefrom acoustic echo and noise as is the case with the third embodiment. Further, according to the fifth embodiment, sensitivity constraints can be imposed on a plurality of sound sources, and sensitivity can be held for a sound source having uttered speech sound previously as well. Accordingly, this embodiment is advantageous in that that even when the sound source moves, the speech quality does not deteriorate in the initial part of the speech sound since the sensitivity for the sound source is maintained if it has uttered speech sound in the past.

SIXTH EMBODIMENT

A sound acquisition apparatus according to a sixth embodiment of the present invention will be described.

In the sound acquisition apparatus of this embodiment, the weighting

factors C_{S1} to C_{SK} of the sensitivity constraints for the sound source positions 9_k in the sound acquisition apparatuses of the first to third and fifth embodiments are changed on a timewise basis.

The time-variant weighting factors C_{S1} to C_{SK} of the sensitivity constraints for the sound sources 9_1 to 9_K are set smaller in order of utterance in the past. A first method is to reduce the weighting factor C_{Sk} with an increase in the elapsed time from the detection of each already detected sound source position to the detection of the most recently detected sound source position. A second method is to set the weighting factor C_{Sk} smaller in order of detection of K sound source positions.

Fig. 14 illustrates in block form the functional configuration of a weighting factor setting part 21H for implementing the above-said first method. The weighting factor setting part 21H is made up of: a clock 21H1 that outputs time; a time storage part 21H2 that upon each detection of sound source position, overwrites the time t of detection, using, as an address, the number k representing the detected sound source 9_k; and a weighting factor determining part 21H3. Based on the time of detection of the sound source position stored in the time storage part 21H2, the weighting factor determining part 21H3 assigns a predetermined value C_S as the weighting factor S_{Ck} to the currently detected sound source of a number k(t), and assigns a value $q^{(t-tk)}C_S$ as the weighting factor C_{Sk} to each of the other sound sources of numbers $k \neq k(t)$ in accordance with the elapsed time $t-t_k$ after the detection time t_k. q is a predetermined value in the range of $0<q\leq 1$. In this way, the weighting factors C_{S1} to C_{SK} of sensitivity constraints are determined for the respective sound sources, and they are provided to 21A1 to 21AK.

Fig. 15 illustrates in block form the functional configuration of a weighting factor setting part 21H for implementing the above-said second

method; in this example, it is made up of a clock 21H1, a time storage part 21H2, an order decision part 21H4, and a weighting factor determining part 21H5. The order decision part 21H4 decides the order of detection of the positions of the sound sources 9_1 to 9_K (the newest order) $\{k(t)\}=\{k(1), \ldots, k(K)\}$ from the times stored in the time storage part 21H2. The weighting factor determining part 21H5 assigns a predetermined value C_S as a weighting factor $C_{Sk(1)}$ to the most recently detected sound source $9_{k(1)}$. For the other sound sources, the weighting factor determining part calculates $C_{Sk(t+1)} \leftarrow qC_{Sk(t)}$ for $t=1,2, \ldots, K-1$ to obtain weighting factors $C_{Sk(2)}, \ldots, C_{Sk(t)}$. These weighting factors $C_{Sk(2)}$ to $C_{Sk(t)}$ are rearranged following the order $\{k(1), \ldots, k(K)\}$, thereafter being output as weighting factors $C_{S1}, \ldots, C_{SK}$. The value of q is a preset value in the range of $0<q<1$.

By varying the weights of sensitivity constrains for the respective sound sources as described above, it is possible to reduce the sensitivity constrains for the sound source positions where utterance was made in the past. Thus, as compared with the sound acquisition apparatuses of the first to third embodiments, the apparatus of this embodiment reduces the number of sound sources to be subjected to sensitivity constraints, enhancing the acquired sound level adjustment capability and the noise and acoustic echo cancellation functions.

The other parts are the same as those in the first to third and fifth embodiments of the present invention, and hence no description will be repeated.

SEVENTH EMBODIMENT

A sound acquisition apparatus according to a seventh embodiment of the present invention will be described.

The sound acquisition apparatus according to the seventh embodiment

of the present invention features whitening the covariance matrix $\mathbf{R}_{XX}(\omega)$ in the filter coefficient calculating part 21 of the sound acquisition apparatus according to the first to sixth embodiments of the present invention. Fig. 16 illustrates the functional configuration of a representative one of whitening parts 21J1 to 21JK indicated by the broken lines in the filter coefficient calculating part 21 shown in Fig. 4. The whitening part 21J comprises a diagonal matrix calculating part 21JA, a weighting part 21JB, an inverse operation part 21JC and a multiplication part 21JD. The diagonal matrix calculating part 21JA generates a diagonal matrix $\mathrm{diag}(\mathbf{R}_{XX}(\omega))$of the covariance matrix $\mathbf{R}_{XX}(\omega)$ fed thereto. The weighting part 21JB assign weights to the diagonal matrix by calculating the following equation based on a matrix $\mathbf{D}$ of a predetermined arbitrary M or M+1 rows.

$$\mathbf{D}^{T}\mathrm{diag}(\mathbf{R}_{XX}(\omega))\mathbf{D} \qquad (34)$$

The inverse calculation part 21JC calculates an inverse of Eq. (34)

$$1/\{\mathbf{D}^{T}\mathrm{diag}(\mathbf{R}_{XX}(\omega))\mathbf{D}\} \qquad (35)$$

In the above ^{T} indicates a transpose of the matrix. In the multiplication part 21JD the result of calculation by the inverse calculation part 21JC is multiplied by each covariance matrix $\mathbf{R}_{XX}(\omega)$ input thereto to obtain a whitened covariance matrix.

With the covariance matrix thus whitened, the filter coefficients obtained in the filter coefficient calculating part 21 no linger change with spectral changes of the send signal, acquired signal and the noise signal. As a result, the acquired sound level adjustment capability and the acoustic echo and noise cancellation capabilities do not change with the spectral changes—this makes it possible to achieve steady acquired sound level adjustment and acoustic echo and noise cancellation.

The other parts are the same as in the first to fourth embodiments of

the present invention, and hence no description will be repeated.

EIGHTH EMBODIMENT

A sound acquisition apparatus according to an eighth embodiment of the present invention will be described.

The sound acquisition apparatus of the eighth embodiment features that the covariance matrix storage part 18 of the sound acquisition apparatus according to the first to seventh embodiments of the present invention averages an already stored covariance matrix and a covariance matrix newly calculated by the covariance matrix calculating part 17 and stores the averaged covariance matrix as the current one.

The covariance matrices are averaged, for example, by the method described below. Letting the already stored covariance matrix be represented by $\mathbf{R}_{XX,old}(\omega)$ and the covariance matrix newly calculated by the covariance matrix calculating part 17 by $\mathbf{R}_{XX,new}(\omega)$, the following equation is used to calculate an average covariance matrix $\mathbf{R}_{XX}(\omega)$.

$$\mathbf{R}_{XX}(\omega)=(1-p)\mathbf{R}_{XX,new}(\omega)+p\mathbf{R}_{XX,old}(\omega) \quad (36)$$

where p is a constant that determines the time constant of the average and takes a value $0\leq p<1$.

Fig. 17 illustrates the functional configurations of the covariance matrix storage part 18 and an averaging part 18A provided therein. The averaging part 18A comprises a multiplier 18A1, an adder 18A2, and a multiplier 18A3. The covariance matrix $\mathbf{R}_{SkSk}(\omega)$ corresponding to the sound source 9_k, calculated by the covariance matrix calculating part 17, is provided as a new covariance matrix $\mathbf{R}_{SkSk,new}(\omega)$ to the multiplier 18A1 and is multiplied by (1−p), and the multiplied output is applied to the adder 18A2. On the other hand, the covariance matrix $\mathbf{R}_{SkSk}(\omega)$ corresponding to the sound source 9_k is read out of the storage area 18B then provided as a old covariance

matrix $\mathbf{R}_{SkSk,old}(\omega)$ to the multiplier 18A3 and multiplied by the constant p. The multiplied output is added by the adder 18A2 to the output $(1-p)\mathbf{R}_{SkSk,new}(\omega)$ from the multiplier 18A1, and the thus obtained average covariance matrix $\mathbf{R}_{SkSk}(\omega)$ is overwritten in the storage area corresponding to the sound source 9_k.

By averaging covariance matrices and storing the averaged covariance matrix as described above, it is possible to lessen the influence of a circuit noise or similar disturbance as compared with that before averaging, hence providing an accurate covariance matrix—this makes it possible to determine filter coefficients that enhance the acquired sound level adjustment, noise cancellation or acoustic echo cancellation performance.

The other parts are the same as in the first to fifth embodiments of the present invention, and hence no description will be repeated.

Incidentally, the present invention can be implemented by dedicated hardware; alternatively, it is possible that a program for implementing the invention is recorded on a computer-readable recording medium and read into a computer for execution. The computer-readable recording medium refers to a storage device such as a floppy disk, an magneto-optical disk, CD-ROM, DVD-ROM, a nonvolatile semiconductor memory, or an internal or external hard disk. The computer-readable recording medium also includes a medium that dynamically holds a program for a short period of time (a transmission medium or transmission wave) as in the case of transmitting the program via the Internet, and a medium that holds the program for a fixed period of time, such as a volatile memory in the computer system serving as a server in that case.

EFFECT OF THE INVENTION

Next, to demonstrate the effectiveness of the first embodiment of the sound acquisition apparatus according to the present invention, Figs. 18A and 18B show the results of simulations with microphones disposed at corners of a square measuring 20 cm by 20 cm. The simulation conditions are—number of microphones: 4, signal-to-noise ratio: about 20 dB, room reverberation time: 300 ms, and number of speakers: 2 (speaker A at a position 50 cm away from the center of the square in a direction at right angles to one side thereof, speaker B at a position 200 cm away from the center of the square in a direction at 90° to the speaker A). Fig. 18A shows microphone received signal waveforms obtained when the speakers A and B spoke alternately under the above-mentioned conditions. Comparison between the speech waveforms of the speakers A and B indicates that the speech waveform of the speaker B is small in amplitude. Fig. 18B shows waveforms processed by the present invention. The speech waveforms of the speakers A and B are nearly equal in amplitude, from which the effect of acquired sound level adjustment can be confirmed.

Fig. 19 shows simulation results obtained with the third embodiment shown in Fig. 10. The simulation conditions are—number M of microphones: 4, signal-to-noise ratio of send signal before processed: 20 dB, send signal-to-acoustic echo ratio: −10 dB, and room reverberation time: 300 msec. Fig. 19 shows the send signal levels obtained when signal sending and receiving are repeated alternately under the above-mentioned conditions. Row A shows the send signal level before processing, and Row B the send signal level after processing by the third embodiment. The above results indicate that the third embodiment reduces the acoustic echo about 40 dB and the noise signal about 15 dB, from which it can be confirmed that the

embodiment of the invention is effective.

As described above, according to the first embodiment of the present invention, it is possible to obtain a send signal of a volume adjusted for each sound source position by: detecting the sound source position from signals picked up by a plurality of microphones; calculating filter coefficients based on a covariance matrix in the utterance period for each sound source position; filtering the microphone acquired signals by the filter coefficients; and adding the filtered signals.

According to the second embodiment of the present invention, it is possible to achieve noise cancellation as well as the acquired sound level adjustment by determining the filter coefficients by using a covariance matrix in the noise period in addition to the covariance in the utterance period in the first embodiment.

According to the third embodiment of the present invention, it is possible to achieve acoustic cancellation by determining the filter coefficients by using a covariance matrix in the receiving period in addition to the covariance matrix in the utterance period in the first or second embodiment.

According to the fourth embodiment of the present invention, it is possible to reproduce the received signal by a loudspeaker and cancel acoustic echo by determining the filter coefficients by using the covariance matrix in the utterance period and the covariance matrix in the receiving period.

According to the fifth embodiment of the present invention, it is possible to further cancel noise by determining the filter coefficients by using the covariance matrix in the noise period in addition to the covariance matrices in the utterance and receiving periods in the fourth embodiment.

According to the sixth embodiment of the present invention, it is possible to further enhance the acquired sound level adjustment, noise

cancellation or acoustic echo cancellation performance by assigning a smaller weighting factor to the covariance matrix of older utterance at the time of calculating the filter coefficients in the first, second, third and fifth embodiments.

According to the seventh embodiment of the present invention, it is possible to implement acquired sound level adjustment, noise cancellation and acoustic echo cancellation not so susceptible to signal spectral changes by whitening the covariance matrix at the time of calculating the filter coefficients in the first to sixth embodiment.

According to the eighth embodiment of the present invention, when the covariance matrix is stored in the first to seventh embodiments, the covariance matrix and that already stored in the corresponding area are averaged and a weighted mean covariance matrix is stored, by which it is possible to obtain a more accurate covariance matrix and determine filter coefficients that provide increased performance in the acquired sound level adjustment, noise reduction and acoustic echo cancellation.